DOCUMENT RESUME

ED 053 214 TM 000 751

AUTHOR Golden, George M.; And Others
TITLE Item Writing Rule Conformity as Related to Biographical Item Response Stability.
PUB DATE May 71
NOTE 6p.; Paper presented in a session on Testing and Training Materials at the Convention of the Rocky Mountain Psychological Association, Denver, Colorado, May 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Academic Achievement, Background, *Biographical Inventories, College Freshmen, *Item Analysis, Placement, *Prediction, *Questionnaires, Self Evaluation, *Test Construction, Test Reliability, Test Validity

ABSTRACT

This study examined relationships among conformity to biographical item writing rules, rated stability, and obtained response stability. First and second administration data were collected and a stability index was obtained on 88 items using a college freshmen sample (N=106). One group of judges then educed rules to discriminate between the more stable and the less stable items, and a second group judged the conformity of each item to the seven educed rules. Intercorrelations were computed among the rule conformity indices, Probable Response Stability scale ratings, and the obtained stability indices. Although three of the rules were significantly related to item stability, using them to compute multiple correlations did not enhance prediction of stability compared with application of the ratings alone. Practical implications for optimizing biographical item stability and improving a priori prediction of response stability in future research are discussed. (Author/AG)

ITEM WRITING RULE CONFORMITY AS RELATED TO BIOGRAPHICAL ITEM RESPONSE STABILITY

GEORGE M. GOLDEN and ALLAN R. STARRY
Measurement and Research Center
Purdue University - AGA 1
Lafayette, Indiana 47907

and I. van W. RAUBENHEIMER
University of South Africa, Pretoria

Paper presented in a session on Testing and Training Materials at the 1971 Convention of the Rocky Mountain Psychological Association in Denver, May 1971

ITEM WRITING RULE CONFORMITY AS RELATED TO BIOGRAPHICAL ITEM RESPONSE STABILITY

George M. Golden and Allan R. Starry
Measurement and Research Center, Purdue University

and I. van Raubenheimer
University of South Africa, Pretoria

Abstract

Widespread use of life history items as predictors of academic and work criteria indicates that additional information should be obtained on the stability of responses to biographical questionnaires. Responses to personal history items are implicitly assumed to remain stable over time in instances where questionnaire responses are weighted and used as a selection or placement tool. Guidelines or techniques aimed at upgrading the art of writing effective life history items have been suggested by many investigators as a way of improving and predicting response stability.

This study examined relationships among conformity to biographical item writing rules, rated stability, and obtained response stability. First and second administration data were collected and a stability index was obtained on 88 items using a college freshman sample (N=106). One group of judges then educed rules to discriminate between the more stable and the less stable items, and a second group judged the conformity of each item to the seven educed rules. Intercorrelations were computed among the rule conformity indices, Probable Response Stability scale ratings, and the obtained stability indices. Although three of the rules were significantly related to item stability, using them to compute multiple correlations did not enhance prediction of stability compared with application of the ratings alone.

Practical implications for optimizing biographical item stability and improving a priori prediction of response stability in future research are discussed.

Introduction

Guidelines or techniques aimed at upgrading the art of writing effective life history items have been suggested by several investigators. Among those reported are item writing rules or classification systems (Owens, Glennon, & Albright, 1962; Larsen, Swarthout, & Wickert, 1967; Starry, 1968; Smart, 1968) and rating scales (Starry, Raubenheimer, & Tesser, 1969). Dependent variables used in these studies have included item objectionability, fakeability, and retest reliability. The present investigation concerned the relationship of item writing rules to retest reliability. It was felt that additional relevant rules should be sought by analyzing an independent sample of items, and that interrelationships among resultant stability predictors or classifiers should be explored.

Procedure

From a 671 item biographical questionnaire administered to several thousand college freshmen, 88 five-alternative, continuously scored items representative of content areas sampled by the complete questionnaire were selected. These items were then readministered 16-20 weeks later to a volunteer group of 106 male and female freshmen. A stability index was obtained for each item by computing the absolute difference between responses given by each subject on the two administrations, summing across all subjects, and dividing by the number responding to that item on both administrations (N per item varied from 104 to 106). This quantity was then subtracted from a constant of 9.00 for all items. The obtained stability index ranged from 8.89 for the most stable item to 7.86 for the least stable.

The 15 most stable and 15 least stable of the 88 items were presented to a group of nine graduate students in psychology with instructions to educe rules which differentiated the more stable from the less stable set. Collation resulted in seven fairly distinct item writing characteristics or rules, as listed below:

1. Factual, verifiable, census-type data should be requested.
2. The information requested should be easily recallable.
3. Self-evaluation should not be required.
4. Responses highly subject to temporal changes in attitude or frame of reference should be avoided.
5. Evaluations or judgments concerning other people should not be required.
6. Responses should not be requested which may rely on a specific situation existing at the time a response is made.
7. Information concerning personal accomplishments should not be requested.

An independent group of nine undergraduate students were then asked to indicate whether each of the original 88 items conformed or did not conform to each characteristic. The total number of judges who indicated that a characteristic applied to an item was regarded as the index of conformity for that item/characteristic combination. Thus an item could have a conformity index ranging from 0 to 9 for each characteristic.

Results

The reliability of the mean number of rules judged as conforming to items was .91 (Winer, 1962, p. 130).

Product-moment correlations (reported in Table 1) were computed across the 88 items among the conformity indices of the seven rules, the item stability index, and the Probable Response Stability ratings which had previously been applied to these items (Starry et al., 1969).

Only three of the rules (numbers one, three and six) were significantly correlated with stability at the .01 level, with the largest coefficient being .39. The multiple correlation of the three significant rules with stability was .40, due to the magnitude of the interrelationships among independent variables. Adding the Probable Response Stability ratings to this battery resulted in a coefficient of .48, as compared to .47 using these ratings alone.

Discussion

Of the rules significantly related to item stability, number three, "Self-evaluation should not be required," was also suggested in a similar form by the Larsen et al. (1967) judges. These investigators found that items calling for self-evaluations or ratings tended to be more fakeable than other item types. It should be pointed out that attitudinal items were not defined as being self-evaluative in either study.

Rule number six corresponds very closely to a rule hypothesized in the Owens et al. (1962) study, "A currently correct response should not be subject to too rapid short term evolution." However, in their sample, items conforming to this rule were no more reliable than those failing to conform.

Keating, Paterson, and Stone (1950) found the type of question which would result through strict adherence to rule number one, requiring each item to be factual and verifiable, to be highly reliable. Unfortunately, such a constraint would reduce biographical questionnaires to classical application blanks and surely restrict their validities. Rather than stating this rule as a condition which items should meet, perhaps it would be better to consider it as an "ideal" form (in terms of maximizing reliability) against which items may be compared.

Interrelationships among these rules and those which have been suggested by others are probably such that the problem of linear restraints would be a consideration in attempts to use a number of rules in combination. For example, the significant rules in this study and the rules requiring (1) that items carry neutral or pleasant connotations for the respondent and (2) that numbers should be used to graduate and define alternatives (Owens et al., 1962) would appear to overlap to a considerable extent. Their combined use in item construction might not add appreciably to the stability of a questionnaire over the use of two or three of these rules, although contingencies would be more clearly defined for the user.

Independence of the particular item set used in this study accounted for the non-replication of the Owens et al. (1962) rule that escape options should be used whenever necessary to insure an appropriate response for each individual (none of our 88 items lacked an appropriate response), and may have been the reason that "item brevity is desirable" was suggested by their, but not our, judges. It would seem that both these rules are clearly in the "basic" category and should routinely be employed in the development of sound items. Other rules with similar rational appeal which have not been mentioned here (no original sources will be attempted) include: "Items used with heterogeneous respondent populations should be free, for the most part, from contamination by age and experience factors," "Items should not require finer discriminations than the respondent can handle," "Purely escape-type options should constitute an independent alternative physically separated from the other options," and "Items instructing respondents to check as many options as apply should be avoided."

There was some evidence in this study that item properties included in such rules constitute the type of benchmarks raters are using when judging items on the Probable Response Stability scale (Starry et al., 1969). Were it not for the low percentage of variance accounted for by this rated stability index (about 25%) against a stability criterion, there would be little point in searching for additional item writing rules. Perhaps the item samples explored to date with the Probable Response Stability scale have been nonrepresentative in terms of the incidence of mechanical flaws which would affect stability, in which case a more satisfying proportion of stability variance may actually be predictable in the unedited population. On the other hand, there may be rules not specifically investigated in this study which would tap important sources of variation unaccounted for by the ratings alone, making the whole matter of a priori prediction of stability more worthwhile.

Influences of stability predictors on validity are as yet largely unknown. However, self-evaluation items have demonstrated validity on a variety of criteria, as have ones requesting the respondent to check as many alternatives as apply to him. It could well be that within these two item types there are discriminable differences which result in some being relatively reliable and others not. Retest and criterion data on a large pool of unedited biographical items could throw considerable light on these and other problems.

References

Keating, E., Paterson, D. G., and Stone, C. H. Validity of work histories obtained by interview. Journal of Applied Psychology, 1950, 34, 1-5.

Larsen, R. H., Swarthout, D. M., and Wickert, F. R. Objectionability and fakeability of biographical inventory items. Paper presented at the annual meetings of the Midwestern Psychological Association, Chicago, May 1967.

Owens, W. A., Jr., Glennon, J. R., and Albright, L. E. Retest consistency and the writing of life history items: A first step. Journal of Applied Psychology, 1962, 46, 329-331.

Smart, B. D. Reducing offensiveness of biographical items in personnel selection: A first step. Studies in Higher Education, 1968, 95, 15-21.

Starry, A.R. The utility of biographical data for screening research and development engineering applicants. Symposium paper presented at the seventy-sixth annual meetings of the American Psychological Association, San Francisco, September, 1968.

Starry, A. R., Raubenheimer, I. van W., and Tesser, A. Stability ratings as classifiers of life history item retest reliability. Journal of Applied Psychology, 1969, 53, 14-18.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Table 1

Correlations Among Conformity Indices

of Item Writing Rules, Probable Response Stability Ratings,

and Obtained Stability Indices

Variable	1	2	3	4	5	6	7	8	9
1 - Conformity Index of Rule 1	--	43	74	59	39	30	-02	73	39**
2 - Conformity Index of Rule 2		--	32	12	24	13	-16	29	17
3 - Conformity Index of Rule 3			--	46	08	64	11	68	32**
4 - Conformity Index of Rule 4				--	27	41	15	65	14
5 - Conformity Index of Rule 5					--	08	-10	31	11
6 - Conformity Index of Rule 6						--	-05	61	23**
7 - Conformity Index of Rule 7							--	05	07
8 - Probable Response Stability Rating								--	47**
9 - Retest Stability Index									--

Note -- Decimals omitted.
** $p < .01$